Bayesian Updating: Continuous Priors

18.05 Spring 2014

Jeremy Orloff and Jonathan Bloom

Problem 6a on pset 5

- Mrs S is found stabbed in her family garden.
- Mr S behaves strangely after her death and is considered as a suspect.
- Investigation shows that Mr S had beaten up his wife on at least nine previous occasions.
- The prosecution advances this data as evidence in favor of the hypothesis that Mr S is guilty of the murder.
- Mr S's highly paid lawyer say, '*statistically*, only one in a thousand wife-beaters actually goes on to murder his wife. So the wife-beating is not strong evidence at all. In fact, given the wife beating evidence alone, it's extremely *unlikely* that he would be the murderer of his wife – only a $1/1000$ chance. You should therefore find him innocent.'

What do you think of the lawyers argument?

Problem 6b on pset 5

In 1999 in Great Britain, Sally Clark was convicted of murdering her two sons after each child died weeks after birth (the first in 1996, the second in 1998). Her conviction was largely based on the testimony of the pediatrician Professor Sir Roy Meadow. He claimed that, for an affluent non-smoking family like the Clarks, the probability of a single cot death (SIDS) was 1 in 8543, so the probability of two cot deaths in the same family was around "1 in 73 million." Given that there are around 700,000 live births in Britain each year, Meadow argued that a double cot death would be expected to occur once every hundred years. Finally, he reasoned that given this vanishingly small rate, the far more likely scenario is that Sally Clark murdered her children.

Continuous range of hypotheses

Example. Bernoulli with unknown probability of success p.
Can hypothesize that p takes any value in $[0, 1]$.
Model: 'bent coin' with probability p of heads.

Example. Waiting time $X \sim \exp(\lambda)$ with unknown λ.
Can hypothesize any $\lambda > 0$.

Example. Have normal random variable with unknown μ and σ. Can hypothesisze (μ, σ) anywhere in $(-\infty, \infty) \times [0, \infty)$.

Review of pdf and probability

X random variable with pdf $f(x)$.
$f(x)$ is a *density*, units: probability/units of x.
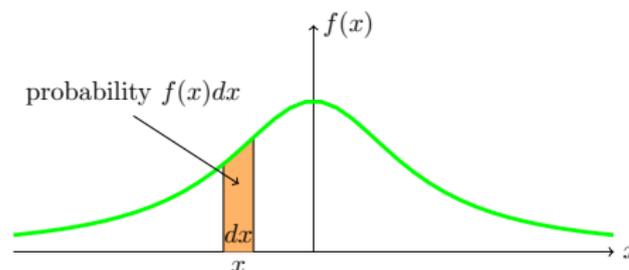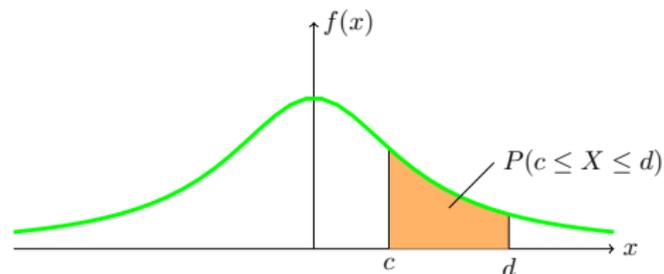
$$P(c \leq X \leq d) = \int_c^d f(x)\,dx.$$

Probability X is in an infitesimal range dx around x is

$$f(x)\,dx$$

Often use probability $f(x)\,dx$ instead of density $f(x)$

Notational clarity for continuous parameters

Example. Suppose that $X \sim$ Bernoulli(θ), where θ is an unknown parameter. We can hypothesize that θ takes any value in $[0, 1]$.

- We use θ instead of p because it's neutral with no confusing connotations.

- Since θ is continuous we need a prior pdf $f(\theta)$.

- Use $f(\theta) \, d\theta$ to work with probabilities instead of densities, e.g. The prior probability that θ is in the range $.5 \pm \frac{d\theta}{2}$ is $f(.5) \, d\theta$.

- To avoid cumbersome language we will say
  'The hypothesis $\theta \pm \frac{d\theta}{2}$ has prior probability $f(\theta) \, d\theta$.'

Concept question

Suppose $X \sim \text{Bernoulli}(\theta)$ where the value of θ is unknown. If we use Bayesian methods to make probabilistic statements about θ then which of the following is true?

1. The random variable is discrete, the space of hypotheses is discrete.

2. The random variable is discrete, the space of hypotheses is continuous.

3. The random variable is continuous, the space of hypotheses is discrete.

4. The random variable is continuous, the space of hypotheses is continuous.

Bayesian update tables: discrete priors

Discrete hypotheses: A, B, C.

Data: $\mathcal{D}$.

Prior probability function: $P(A)$, $P(B)$, $P(C)$.

hypothesis	prior	likelihood	unnormalized posterior	posterior
A	$P(A)$	$P(\mathcal{D}|A)$	$P(\mathcal{D}|A)P(A)$	$\dfrac{P(D|A)P(A)}{P(\mathcal{D})}$
B	$P(B)$	$P(\mathcal{D}|B)$	$P(\mathcal{D}|B)P(B)$	$\dfrac{P(D|B)P(B)}{P(\mathcal{D})}$
C	$P(C)$	$P(\mathcal{D}|C)$	$P(\mathcal{D}|C)P(C)$	$\dfrac{P(D|C)P(C)}{P(\mathcal{D})}$
Total	1		$P(\mathcal{D}) = \text{sum}$	1

Note $T = P(\mathcal{D}) = $ the prior predictive probability of $\mathcal{D}$.

Bayesian update tables: continuous priors

$X \sim$ Bernoulli(θ). Unknown θ

Continuous hypotheses θ in [0,1].
Data x.
Prior pdf $f(\theta)$.
Likelihood $p(x|\theta)$.

hypoth.	prior	likel.	un. posterior	posterior
$\theta \pm \frac{d\theta}{2}$	$f(\theta)\,d\theta$	$p(x\|\theta)$	$p(x\|\theta)f(\theta)\,d\theta$	$\dfrac{p(x\|\theta)f(\theta)\,d\theta}{T}$
Total	1		$T = \displaystyle\int_0^1 p(x\|\theta)f(\theta)\,d\theta$	1

Note $T = p(x)$, the prior predictive probability of x.

Example

'Bent' coin: unknown probability θ of heads.

Flat prior: $f(\theta) = 1$ on $[0, 1]$.

Data: toss once and get heads.

hypoth.	prior	likelihood	unnormalized posterior	posterior
$\theta \pm \frac{d\theta}{2}$	$d\theta$	θ	$\theta\, d\theta$	$2\theta\, d\theta$
Total	1		$\int_0^1 \theta\, d\theta = 1/2$	1

Posterior pdf $= f(\theta|x) = 2\theta$. (Should graph this.)

Board question

'Bent' coin: unknown probability θ of heads.

Prior: $f(\theta) = 2\theta$ on $[0, 1]$. (This is the posterior from the last example.)

Data: toss (again) and get heads.

1. Find the posterior pdf to this new data.

2. Suppose you toss again and get tails. Update your posterior from problem 1 using this data.

3. On one set of axes graph the prior and the posteriors from problems 1 and 2.

Board Question

Same scenario: bent coin $\sim$ Bernoulli(θ).

Flat prior: $f(\theta) = 1$ on $[0, 1]$

Data: toss 27 times and get 15 heads and 12 tails.

1. Use this data to find the posterior pdf.

Give the integral for the normalizing factor, but do not compute it out. Call its value T and give the posterior pdf in terms of T.

Beta distribution

$Beta(a, b)$ has density

$$f(\theta) = \frac{(a + b - 1)!}{(a - 1)!(b - 1)!} \theta^{a-1}(1 - \theta)^{b-1}$$

http://ocw.mit.edu/ans7870/18/18.05/s14/applets/beta-jmo.html

Observation: The coefficient is a normalizing factor so if

$$f(\theta) = c\theta^{a-1}(1 - \theta)^{b-1}$$

is a pdf, then

$$c = \frac{(a + b - 1)!}{(a - 1)!(b - 1)!}$$

and $f(\theta)$ is the pdf of a Beta(a, b) distribution.

18.05 Introduction to Probability and Statistics

Spring 2014